WeData's Ultimate R Book

Vestin Hategekimana

2024-01-21

Table of contents

Preface

The aim of this book is to be a compendium of knowledge on R made by members of the WeData association. It should serve both as a reference for knowledge, and as an exercise interface. We hope to highlight some important themes in this book.

Why R?

- It is free
- Data analysis and statics ecosystem
- Beautiful visualization capabilities
- Amazing reporting tools
- Web application development
- Vibrant and nice community
- Widely used in research
- Key tools in specific sectors

Why this book?

Why this book? To complete existing resources with exercises. In fact, there are already numerous free or paid resources on R scattered on the internet. The main concern is not about knowing if a resource exists but about finding it. Rarely do collaborative projects try to bring together various R resources in the form of a book or website. Here are some notable examples I know of (there are probably more):

- The big book of R
- Bookdown
- Pyofile (R and Python books)

It is amusing to find excellent resources hidden in Rmarkdown and Quarto galleries. We believe that despite the dissemination of resources, the community is doing a fantastic job trying to consolidate them, and I strongly encourage anyone interested in contributing to this effort to feed the sources mentioned earlier. However, another problem remains: there is a glaring lack of exercises and interactive resources. It's important to note that there are platforms for doing R exercises online to some degree of interactivity. Notable cases include Exercism, Codewars and W3school (for a more comprehensive list, check our blog post). There is also the learnr package, which allows creating interactive R exercises using Rmarkdown. The system is great (at least locally), but when it comes to putting exercises online for use without installing R, it requires an active server, adding more work and concerns, particularly regarding hosting and managing a large number of requests. Our goal would be to create an exercise platform that does not require running R on a server. Is this possible? The answer is yes!

The advancements in web technologies over the past few years have given us hope to create online interactive exercise platforms without a server, and it was only recently that the idea of reusing a system like learnr within webr (the Quarto extension) surfaced. These developments are very encouraging and have inspired the creation of this platform.

Our association is made up of passionate individuals with varying levels, goals, and interests. But I believe we can create an interactive resource that allows curious people to train. The objective will be to create exercise series and recommend existing resources to help users complete the exercises.

This is new territory for me, and I don't know if it will work. Perhaps I'm missing essential

elements, or web pages might be unbelievably slow or inefficient. But I believe it's worth a try!

What you will learn in this book

Coming soon...

Exercice example from quarto-live

[quarto-live](#)

Filter the `starwars` dataset so that only the droid characters are included.

```
#| setup: true
#| exercise: ex_3
library(dplyr)
```

```
#| exercise: ex_3
starwars |> _____
```

> **i Hint 1**
>
> Consider using the `filter()` function from `dplyr`.
>
> ```
> starwars |> filter(_____)
> ```

> **i Hint 2**
>
> You should filter the dataset using the `species` column.
>
> ```
> starwars |> filter(species == _____)
> ```

Solution.

> 💡 Fully worked solution:
>
> Use the `filter()` function from `dplyr`:
>
> ```r
> starwars |>                          ①
>     filter(species == "Droid")       ②
> ```
>
> ① Take the `starwars` dataset, and then,
> ② Filter for the "Droid" species.

Part I

Base R

1 Introduction

This is a book created from markdown and executable code.

See Knuth (1984) for additional discussion of literate programming.

```
1 + 1
```

```
[1] 2
```

2 Summary

In summary, this book has no content whatsoever.

```r
1 + 1
```

```
[1] 2
```

References

Knuth, Donald E. 1984. "Literate Programming." *Comput. J.* 27 (2): 97–111. https://doi.org/10.1093/comjnl/27.2.97.

Part II

Math with R

Part III

Tidyverse

Part IV

Statistical analysis

Part V

Machine learning

Part VI

Text mining/NLP

Part VII

Web scraping